

RESEARCH

Open Access



# Prediction and analysis of Indonesia Presidential election from Twitter using sentiment analysis

Widodo Budiharto\*  and Meiliana Meiliana

\*Correspondence:  
wbudiharto@binus.edu  
Computer Science  
Department, School  
of Computer Science,  
Bina Nusantara University,  
Jakarta 11480, Indonesia

## Abstract

Big data encompasses social networking websites including Twitter as popular micro-blogging social media platform for a political campaign. The explosive Twitter data as a respond of the political campaign can be used to predict the Presidential election as has been conducted to predict the political election in several countries such as US, UK, Spain, and French. The authors use tweets from President Candidates of Indonesia (Jokowi and Prabowo), and tweets from relevant hashtags for sentiment analysis gathered from March to July 2018 to predict Indonesian Presidential election result. The authors make an algorithm and method to count important data, top words and train the model and predict the polarity of the sentiment. The experimental result is produced by using R language and show that Jokowi leads the current election prediction. This prediction result is corresponding to four survey institutes in Indonesia that proved our method had produced reliable prediction results.

**Keywords:** Sentiment analysis, Twitter, Presidential election, Prediction

## Introduction

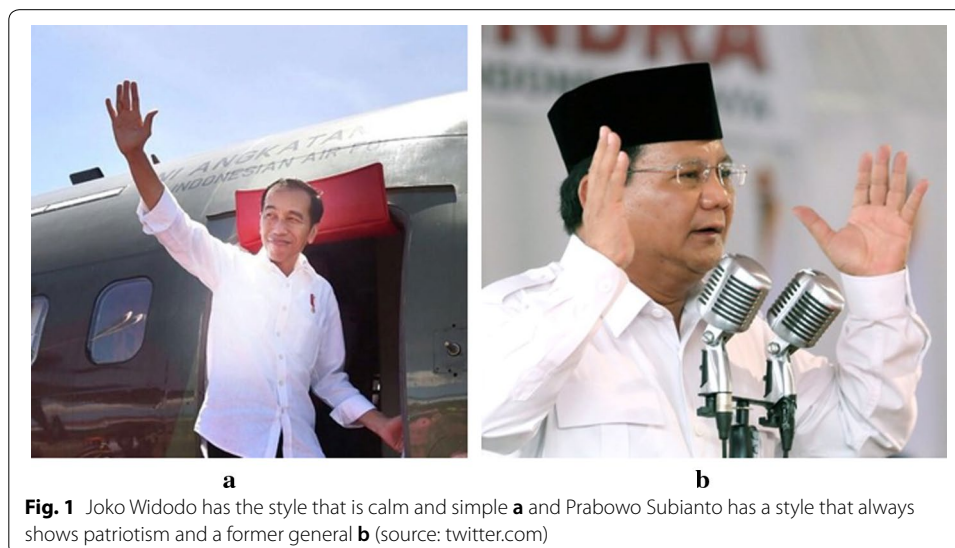
Elections in Indonesia have taken place since 1955 to elect a legislature. At a national level, Indonesian people did not elect a president until 2004. For the first time, the president, and members of People's Consultative Assembly will be elected on the same day [1]. The next general election that will be held in Indonesia is next year on 17 April 2019. Related to this situation, discussion and prediction about who is the Presidential candidate in Indonesia become a hot and interesting conversation among Indonesian citizen, and many of them expressed it through social media. Election-related hashtags are some of the most used hashtags among Indonesian netizens, most of them is a form of support to Jokowi and Prabowo, such as #PilihPrabowo (vote for Prabowo) and #AkhirnyaMilihJokowi (finally vote for Jokowi) [2]. Political campaigns have exploited this vast array of information available on the above platforms to draw insights about user opinions and thus design their campaign strategy. Huge investments by politicians in social media campaigns right before an election along with arguments and debates between their supporters and opponents only enhance the claim that views and opinions posted by users have a bearing on the results of an election [3]. On the other way, the information

provided could be used to predict the election result by using data analysis method such as sentiment analysis.

Jokowi is the incumbent president and challenged by Prabowo who lost in the last Presidential election. The pictures of Jokowi and Prabowo are shown in Fig. 1.

Sentiment analysis is an analysis to identify customer like, dislike, comment, opinion, or feedback about a content that will be categorized into positive, negative or neutral responses. Social media plays a significant role in sentiment analysis. From the survey in 2017, over 143 million Indonesians use the internet, and approximately 90 percent of these people are using Twitter, Facebook or Instagram. Twitter is micro-blogging social networking of textual message. The messages posted through this social media platform are called as Tweets. The tweets itself since September 2017 fit 280 characters for each post and available as public data. Compare to the other two social media platforms that focus on image and could content long text document. Twitter provides more compact and meaningful data to express an opinion. Thus, this research focuses on Twitter data to provide more reliable data for sentiment analysis as part of the prediction method.

Explosive data available online as the result of significant social media usage could be used as data source to predict the political election result. Compare to the conventional way of offline polling, the prediction of the election result by using twitter data is more effective both in cost and time. Some similar researches have been conducted to predict election result in other countries such as United States, United Kingdom, Spain, and French. Each research proposed a different method and approach, but most of them were using Twitter data as the primary tool that has been proved to be valid and effective source [4]. Prediction framework by using Twitter such as proposed by Kalampokis et al. in 2017 comprises two phases namely Data Conditioning phase and Predictive Analysis phase [5]. The data condition phase consists of the determination of time window, identification of location, user profile characteristics and selection of search terms. The predictive analysis phase consists of the computation of predictor variables, the creation of a predictive model and evaluation of the Predictive Performance [6].



This paper proposes a new framework to predict the election result and sentiment analysis from Twitter data that focuses on Indonesia Election in 2019. The organization of this paper is as follows: starting with an introduction about Presidential election using twitter, the second section discusses related work and the subsequent section describes the proposed method. The fourth section presents the experimental results and discussion; the conclusion is given in the last section.

### Related work

Real human languages provide many problems for Natural Language Processing (NLP) such as ambiguity, anaphora, and vagueness. The authors use R languages and many libraries such as *sentiment* that is designed to quickly calculate text polarity sentiment at the sentence level and optionally aggregate by rows or grouping variables [7]. Recent developments in the field of social media such as Twitter and Instagram usually using Open Authorization (OAuth) to access Twitter and we can access data from R using APIs [8].

Abbas et al. test the efficient market hypothesis to see if Twitter aggregates information faster than a real-money prediction market. They use Support Vector Machines (SVMs), a supervised learning algorithm, to predict the outcome of the 2012 US Presidential elections via Twitter data. We then compare the prediction from SVM against the Iowa Electronic Markets (IEM). A total of 40 million unique tweets were collected and analyzed between September 29th, 2012 and November 6th, 2012. The SVM prediction results are positively correlated with the IEM and predict Obama winning the election, implying that Twitter can be considered as a valid source in predicting US Presidential election outcomes [9]. Huyen et al. used the United States election in 2016 as the source data from Twitter. The Twitter mining was not aiming to predict the election result, but rather to provide a rich analysis of online tweet. They measure party, personality and policy impact aspect of crucial candidacy announcement [10]. Hamling et al. also focused on 2016 US election on their research. They wrote a program to collect tweets that mentioned one of the two candidates, then sorted the tweets by state and developed a sentiment algorithm to see which candidate the tweet favored, or if it was neutral [11].

Ibrahim et al. present approach for predicting the results of Indonesia Presidential Election using Twitter as the main resource. First, they collected Twitter data during the campaign period. Second, they performed automatic buzzer detection on Twitter data to remove those tweets generated by computer bots, paid users, and fanatic users that usually become noise in data. Third, they performed a fine-grained political sentiment analysis to partition each tweet into several sub-tweets and subsequently assigned each sub-tweet with one of the candidates and its sentiment polarity. Their study suggests that Twitter can serve as an important resource for any political activity, specifically for predicting the final outcomes of the election itself [12]. Another research was conducted by Wang et al. that predicts the result of the 2017 French Presidential election by extracting and analyzing sentimental information from Twitter. The proposed method by Lei Wang considers neutral tweets related to specific candidates, which has been proved to increase prediction accuracy in our case study of predicting the 2017 French election result [13]. From most of the related research mentioned in this section,

we could conclude how sentiment analysis according to Twitter data was somewhat accurate to predict election result from all around the world. This paper focuses on the 2019 Indonesia Presidential election of Twitter data by using a new proposed framework that combines tweet counting and sentiment analysis as the pre-processing work.

Twitter data of 2015 UK General Election used by Burnap [14] to forecast the election result. Burnap proposed baseline model that incorporates prior party support and sentiment analysis to generate an accurate forecast of parliament seat allocation. Soler [4] developed a tool to define experiments and to capture the defined conversations and have applied it to the cases of three Spanish elections during 2011 and 2012. Soler concludes that Twitter may be a valid tool for predicting election result, confirm several aforementioned researches such as [9, 12].

### **Sentiment analysis in Twitter**

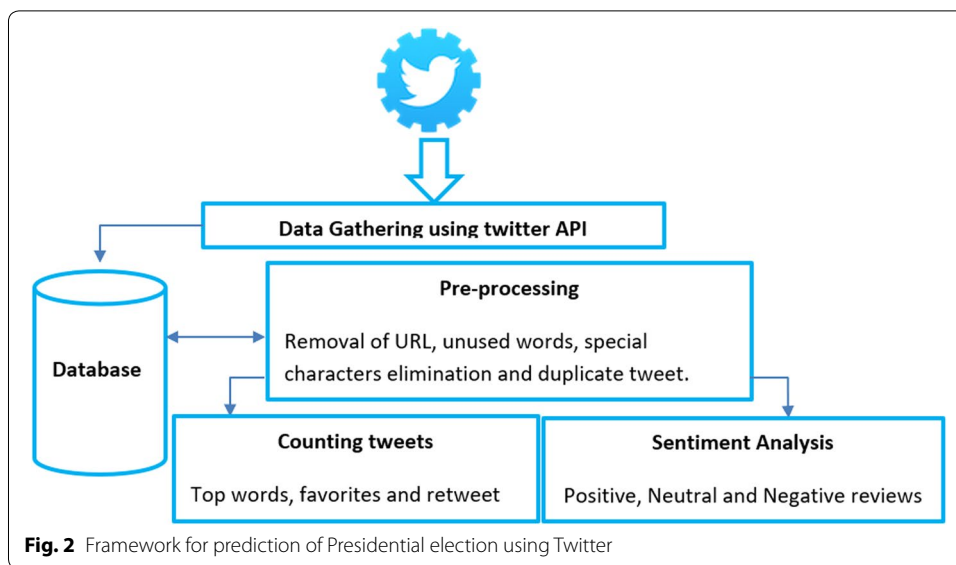
Sentiment analysis, also called opinion mining, is the field of study that analyzes people's opinion, sentiment, evaluation, appraisal, attitude, and emotion towards entities such as product, service, organization, individual, issue, event, topic, and their attributes. The term sentiment analysis introduced in [15] and the term opinion mining is from [16]. Sentiment analysis has been handled as a Natural Language Processing task at many levels of granularity. In the political field, it is used to keep track of political view, to detect consistency and inconsistency between statements and actions at the government level. It can be used to predict election results as well.

Sentiment analysis in Twitter is started by crawling tweets against hashtags to collect all related data. The next step is to do tweets preprocessing and cleaning. Some processes that could be conducted for tweets preprocessing are: removing twitter handles (@user); removing punctuation, numbers and special characters; removing short words; tokenization; and stemming. The cleaned tweets then could be analyzed and visualized based on a specific purpose. Sentiment analysis generally will create or find a list of words associated with strongly positive or negative sentiment. Many positive words and a few negative words indicate positive sentiment, while many negative words and few positive words indicate negative sentiment.

### **Proposed method**

The authors propose the framework that explains the step of the collection, sentiment analysis, and classification of Twitter opinions. Authors have created an account on Twitter API linked to the Twitter account, and Twitter API Authentication process is carried out using OAuth package of R language [17]. Twitter app is used to gather tweets from Jokowi and Prabowo and get the public opinion based on collected hashtags related to views about the Presidential election. To retrieve the tweets, Twitter API accepts parameters and provides the Twitter account's data in return. Retrieved tweets were saved in the database under the following fields such as `twitter_id`, `hashtag`, `tweet_created`, `tweet_text`, `retweet_count`, `favorite_count`. The authors collect Twitter data archives then the process of sentiment analysis is to calculate the synchronization of the words of the tweets with respect to positive, neutral and negative word list.

Figure 2 shows the framework for Presidential election using Twitter. Based on Fig. 1, after the authentication process, data gathered is stored in a database. Pre-processing



**Table 1** Some hashtags related to Presidential election in Indonesia

Hash tags					
#Pilpres2019	#Pemilu2019	#Noblackcampain	#Indonesiabisa	#NKRI	#Indonesiahebat

consists of URL removal, unused words such as stop words in Indonesian language and special characters elimination. After that, we can count tweet to obtain top keywords, favorite lines, and re-tweet. On the sentiment analysis phase, authors calculate the positive, neutral and negative reviews.

The authors select hashtags that were trending on Twitter, representing the political views of people, as shown in Table 1.

For sentiment analysis, the authors use the training set with 250 tweets, and the test set 100 tweets, because the limitation of data. Polarity was calculated using TextBlob. For top keyword, the authors use 5 months data for getting knows the main top keywords for each candidate. The authors use a useful approach to define the score formula as below:

$$\text{Score} = \text{Number of positive words} - \text{Number of negative words} \tag{1}$$

If  $\text{Score} > 0$ , this means that the sentence has an overall 'positive opinion'

If  $\text{Score} < 0$ , this means that the sentence has an overall 'negative opinion'

If  $\text{Score} = 0$ , then the sentence is considered to be a 'neutral opinion'

Polarity gives the differences between the number of positive and the number of negative words in each text, divided by the total number of sentiment words. Authors developed the program using R language that consists of three steps: access the twitter data, preprocessing, count tweet and sentiment analysis. The algorithm for prediction of the Presidential election is shown below:

Algorithm 1. Algorithm for Presidential election based on Twitter based on R language and packages.

```

Begin
  // Access the tweet using our own consumer_key, consume_secret, access_token, and access_secret
  setup_twitter_oauth(consumer_key, consumer_secret, access key, access_secret);
  //Extract tweets from a single user at a time and put the tweets downloaded into a data.frame
  Jokowi_tweets<- userTimeline(user = "@jokowi")
  Prabowo_tweets<- userTimeline(user = "@prabowo")
  Jokowi_tweets<- twListToDF(Jokowi_tweets)
  Prabowo_tweets<- twListToDF(Prabowo_tweets)
  count() // calling function count()
  sentimentanalysis() //calling function sentimentanalysis()
end
function count() //function for calculating tweet's properties from candidates
  for all tweets from Candidates
  do
    retrieve tweets
    removeunused words
    if (tweet_text ==previous_tweet) //if duplicate tweet
      remove duplicate tweet
    end if
    store_to_database(tweet_id, tweet.retweet_count,
tweet.favorite_count,tweet.created_at, tweet.text,tweet.user.id,
tweet.user.followers_count, hashtag)
    count tweets and favorites
  end for
end function
functionsentimentanalysis()
training dataset
for all hashtagfromhashtag_list //processing the tweet
do
  retrieve tweets
  unused words removal
  store_to_database(tweet.retweet_count,
tweet.favorite_count,tweet.id,HashTag,tweet.created_at, tweet.text,tweet.user.id,
tweet.user.followers_count,tweet.user.screen_name)
  calculate polarity
end for

```

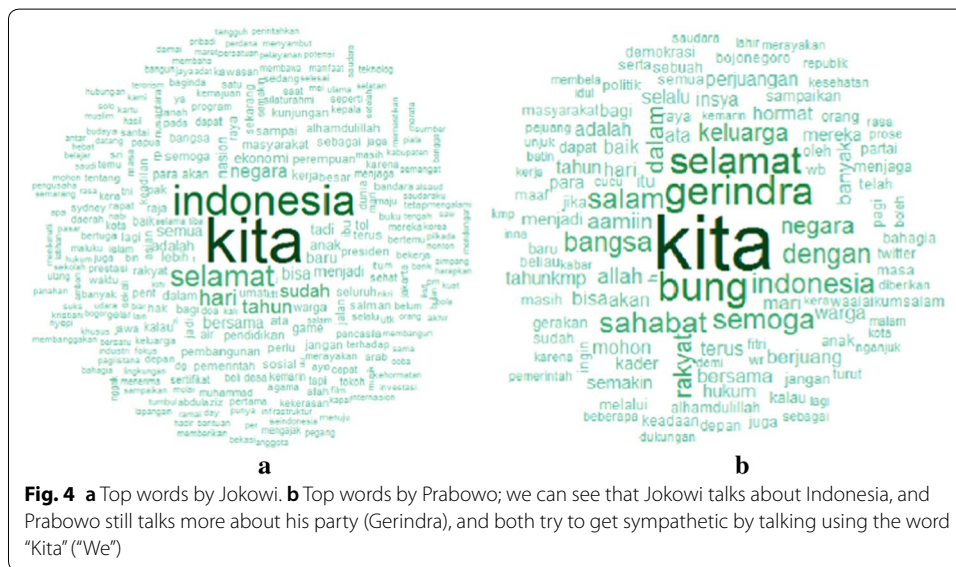
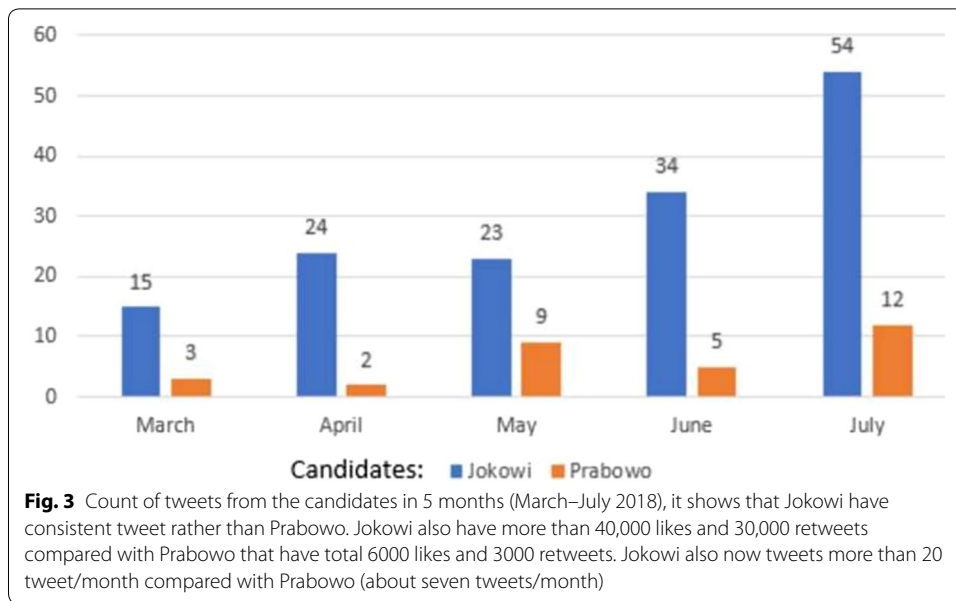
## Experimental result and discussion

We collect Twitter data directly on the web using data from March to July 2018. User @Jokowi has 10.4 M followers, and user @Prabowo has 3.24 M followers based on data retrieved on 16 September 2018. The result of how many tweets from the candidate's account is shown in Fig. 3, and top words by candidates shown in Fig. 4.

We can make a dendrogram as shown in Fig. 5 to shows the most words used by Prabowo in detail:

Based on the data, total likes, followers and retweets for Jokowi are very high compared with Prabowo. The average like for Jokowi's tweets with more than 10 million followers is 9000 and retweets about 3000. The average like for Prabowo's tweets with more than 3000 followers is 1000 and retweets about 500. Figure 6 shows the result of sentiment analysis from the candidates. It seems that Jokowi still has more positive response from the citizen. Unfortunately, Prabowo has more negative sentiment because he has some negative issues about his party and his supporters.

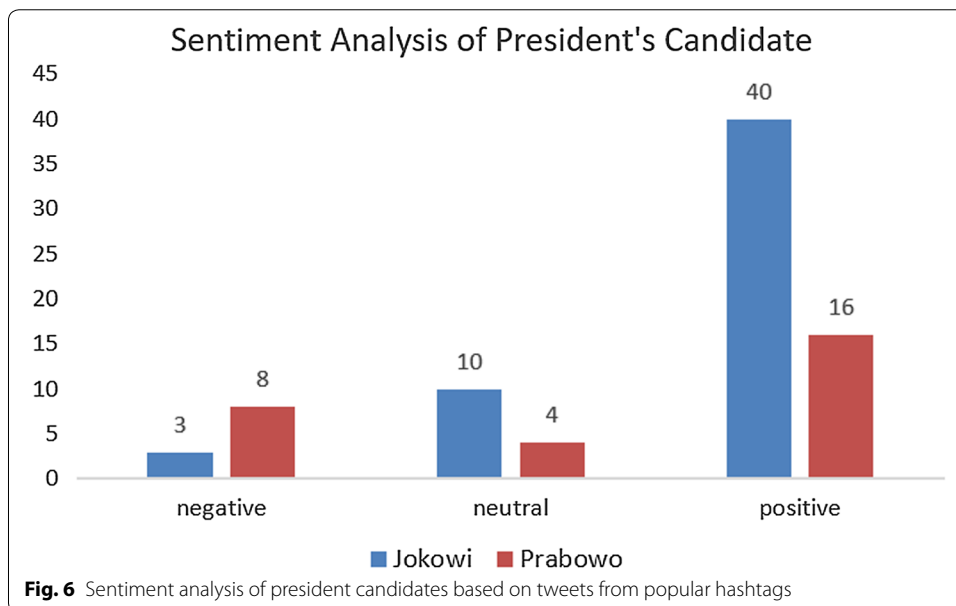
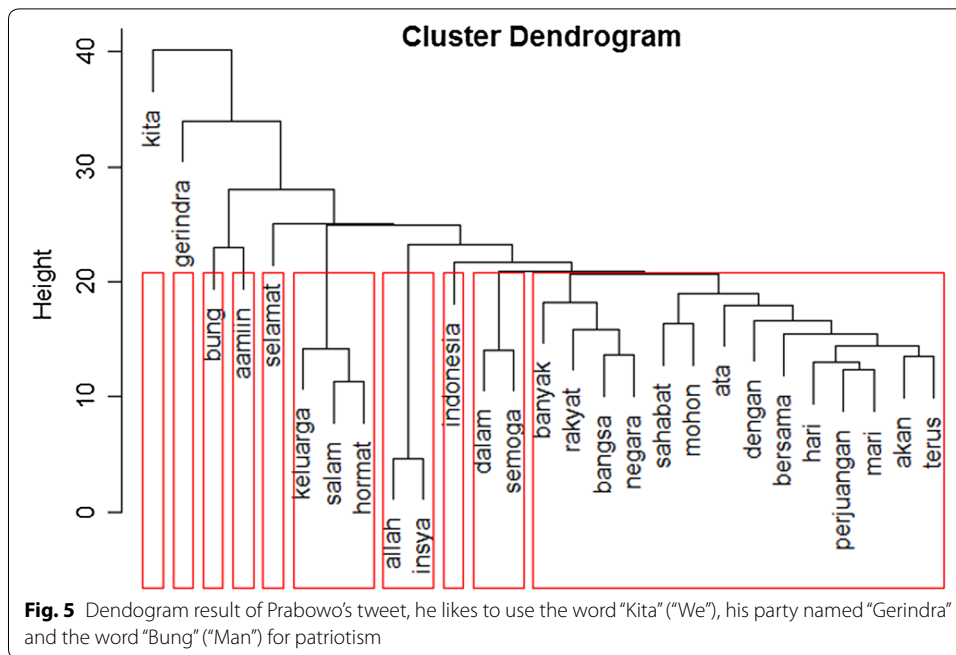




As an additional, twitter is proved to be an essential app in Indonesia, newest information and results show that the candidates that have many likes in tweet and retweet become a winner of district election such as Khofifah Indar Parawansa as a governor of East Java and Ridwan Kamil as a Governor of West Java.

**Conclusions**

Twitter proved to be a valid tool for a poll or opinion mining, especially to predict the outcome of a political election result. Several researches have been conducted to predict election in United States, United Kingdom, Spain, French and Indonesia itself. On this research, the authors focus on tweets data related to 2019 Presidential election with top keywords



that could be seen in Fig. 3. The authors use Twitter data from March 2018, where the discussion about the new election is started to be posted, until July 2018 (the time we conducted the experimental work). Based on those data, the authors proposed a new method to predict the election result that focuses only on tweet counting and sentiment analysis as the preprocessing task. We can easily access tweets of candidates using Twitter API. This method is a way simpler than other methods yet proved to be sufficient to produce a reliable result since both aspects have a significant contribution to the prediction. The experimental result is produced by using R language and show that Jokowi leads the current



election prediction and increasing until this time. This prediction result is corresponding to four survey institutes in Indonesia; Indikator, Cyrus Networks, LitbangKompas and Pol-tracking as mention in Detik News [18]. For the future works, the authors will continue mining and analyzing more Twitter data until around the election time and after the election to get a more accurate prediction.

#### Authors' contributions

Both authors read and approved the final manuscript.

#### Acknowledgements

We say thanks to Bina Nusantara University for supporting this research.

#### Competing interests

The authors declare that they have no competing interests.

#### Availability of data and materials

Not applicable.

#### Consent for publication

Not applicable.

#### Ethics approval and consent to participate

Not applicable.

#### Funding

Not applicable.

#### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 10 July 2018 Accepted: 7 December 2018

Published online: 19 December 2018

#### References

1. Indonesian Presidential election on 17 April 2009, Kompas Newspaper, 26 April 2017. Accessed 3 May 2017.
2. Penetration of Internet in Indonesia. <https://dailysocial.id/post/apjii-survei-internet-indonesia-2017>. Accessed 20 Feb 2018.
3. Jyoti R, Samarth S, Darshan G and Adil S. (2016) Election result prediction using Twitter sentiment analysis. In: International conference on inventive computation technologies (ICICT), India. Piscataway: IEEE; pp. 1–5.
4. Soler JM, Cuartero F, Roblizo M. (2012). Twitter as a tool for predicting elections results. In: IEEE/ACM International Conference on advances in social networks analysis and mining (ASONAM), 2012. Piscataway: IEEE; pp. 1194–1200.
5. Kalampokis E, Karamanou A, Tambouris E, Tarabanis KA. On predicting election results using twitter and linked open data: the case of the UK 2010 election. *J UCS*. 2017;23(3):280–303.
6. Sharma Y, Mangat V, Mandeep K. Sentiment analysis and opinion mining. *Int J Soft Comput Artif Intell*. 2015;3(1):59–62.
7. Kao A, Poteet SR. Natural language processing and text mining. Berlin: Springer; 2007.
8. Ravindran SK, Garg V. Mastering Social media mining with R: extract valuable data from your social media sites and make better business decisions using R. Birmingham: Packt Publisher; 2015.
9. Attarwala A, Dimitrov S, Obeidi A. How efficient is Twitter: Predicting 2012 US Presidential elections using Support Vector Machine via Twitter and comparing against Iowa Electronic Markets. In: Intelligent Systems Conference (IntelliSys), 7–8 September, London; 2017.
10. Le H, Boynton GR, Mejova Y, Shafiq Z, Srinivasan P. Bumps and bruises: mining Presidential campaign announcements on Twitter. In: Proceedings of the 28th ACM conference on hypertext and social media. New York: ACM; pp. 215–224. 2017.
11. Hamling T, Agrawal A. Sentiment analysis of tweets to gain insights into the 2016 US election. *Columbia Undergraduate Sci J*. 2017;11:34–42.
12. Ibrahim M, Abdillah O, Wicaksono AF, Adriani M. Buzzer detection and sentiment analysis for predicting Presidential election results in a Twitter Nation. In: IEEE international conference on data mining workshop. Piscataway: IEEE; pp. 1348–1353. 2015.
13. Wang L, Gan JQ. Prediction of the 2017 French election based on Twitter data analysis. In: Computer Science and Electronic Engineering (CEECE), 2017. Piscataway: IEEE; pp. 89–93. 2017.
14. Burnap P, Gibson R, Sloan L, Southern R, Williams M. 140 characters to victory?: Using Twitter to predict the UK 2015 General Election. *Elect Stud*. 2016;41:230–3.
15. Nasukawa, Tetsuya and Jeonghee Yi, (2003). Sentiment analysis: Capturing favorability using natural language processing. in Proceedings of the 2nd Intl. Conf. on Knowledge Capture. KCAP-03.
16. Kushal D, Lawrence S, Pennock DM. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In: Proceedings of international conference on world wide web (WWW2003). 2003.

17. How to access Twitter data using R language, accessed at <https://medium.com/@GalarnykMichael/accessing-data-from-twitter-api-using-r-part1-b387a1c7d3e>. Accessed 20 Feb 2018.
18. "Tren Elektabilitas Jokowi vs Prabowo di 4 Lembaga Survey" on 4 May 2018. <https://news.detik.com/berita/4003838/tren-elektabilitas-jokowi-vs-prabowo-di-4-lembaga-survei>. Accessed 31 May 2018.

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

---

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)

---